DOCUMENT RESUME

ED 257 631 SE 045 567

AUTHOR Kalomitsines, Spyros
TITLE Some New Ways of Proceeding in Problem Solving.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
REPORT NO LRDC-1985/2
PUB DATE 85
NOTE 52p.
PUB TYPE Guides - Classroom Use - Guides (For Teachers) (052)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Algebra; Computer Simulation; *Computer Software; Geometric Concepts; *Mathematics Instruction; *Problem Solving; Secondary Education; *Secondary School Mathematics; *Teaching Methods

ABSTRACT

Two methods for solving problems are presented, designed to help people overcome difficulties that often occur in the problem-solving process. One difficulty people have is in using their knowledge to reach the decisive solution idea. The proposed "method of description" for eliminating this difficulty is based on a directed enrichment of the problem space. It is presented both as it has been used with secondary school students and as a list of principles. A condensed computer program is also presented; the program automatically solves rather difficult problems by using this method in a small area of algebra and geometry. A second difficulty people have in following an alternative procedure to solve a problem is that they make circles around the problem. The method of "getting out of loops," used to generate new approaches to the problem, is described. (Author/MNS)

LEARNING RESEARCH AND DEVELOPMENT CENTER

1985/2

SOME NEW WAYS OF PROCEEDING IN PROBLEM SOLVING

SPYROS KALOMITSINES

University of Pittsburgh

SOME NEW WAYS OF PROCEEDING IN PROBLEM SOLVING

Spyros Kalomitsines
Experimental School of The University of Athens

Learning Research and Development Center
University of Pittsburgh

1985

The research reported herein was carried out at the Learning Research and Development Center during the period S. Kalomitsines was a Visiting Scholar.

Foreword

Spyros Kalomitsines is a teacher of problem solving who writes from his experience in training students with strong desires to succeed in solving difficult problems in mathematics. This report describes general methods of problem solving that he has found to be helpful in his teaching. It also presents the results of a theoretical effort in which he formulated the heuristic principles of one of the methods explicitly, and implemented them in a computational model of problem solving.

Many of us hope that the concepts and methods of cognitive psychology can be used profitably by teachers and other developers of educational materials to help provide clear and definite expressions of their ideas and methods. Kalomitsines' report gives an example that provides justification for this hope.

J. G. Greeno

Abstract

This study concerns two main difficulties experts and novices often have in solving problems, and suggests approaches for eliminating them. The first is the difficulty people have in using their knowledge to reach the decisive solution idea. The proposed method for eliminating this difficulty consists of a directed enrichment of the problem space, and will be presented in the way it is used in class, as well as in a different form, as a problem-solving process. A computer program which solves problems by using the method is also presented. A second difficulty people have in following an alternative procedure is that they make circles around the problem. The method proposed here presents a scheme to be used in generating new approaches to the problem.

SOME NEW WAYS OF PROCEEDING IN PROBLEM SOLVING

Two new methods will be presented here, to help people overcome difficulties that often occur in problem solving.

The first method is called the "method of description" and is used when the problem solver has come to an impasse in his solution attempt. The method is based on a directed enrichment of the problem space, and will be presented here in two forms: First, the original form of the method will be presented, as it has been used for about three years in the author's classes of 15 to 18 year-old high-school students. Second, a more formal version of the method will be presented, as a list of principles. Finally, a condensed computer program will be presented, where the program automatically solves rather difficult problems by using the method in a small area of algebra and geometry.[1]

The second method, called "getting out of loops," is used when, in the solution process, we see that we are looping in circles around the problem. This method helps one to get out of the loop, by using a novel scheme that generates new approaches to the problem.

---

[1]This list of principles was made at the suggestion of James Greeno.

The Method of Description

In order to make their students more efficient in solving problems, many mathematics teachers teach special methods (heuristic strategies). One such method is used for solving complex problems with many variables. In this method a person solves an analogous problem with fewer variables and then tries to exploit either the method or the result of that solution (Schoenfeld, 1980). Another such method is the pattern of two loci (Polya, 1967). Many lists of such useful instructions are available.

Experienced problem solvers in a specific domain (e.g., algebra) can, after careful observation, classify solutions according to the common characteristics they have, so they can derive useful heuristic strategies for other similar problems. But as every problem solver knows, this procedure of making heuristic strategies for special cases of problems is endless, because there will always be problems where none of these strategies would work. We should always try to extend the list of these strategies to new areas. But it may not be possible to devise a new heuristic strategy before finding the solution of a novel problem. Such a thing might even be a contradiction. We can only try possible strategies that worked for other partially similar problems, or we can apply theorems, etc., hoping that we can find the solution.

Sometimes we can, after a few trials, decide roughly what the solution might be like and develop a strategy that seems to be suitable in this situation. In this way, new strategies can be found while working on a problem, but invention of new strategies before any trial seems very unlikely.

A common situation is one where we have reached an impasse, either from the beginning, or after some steps, although the required knowledge exists in our minds. Let me emphasize here, that I believe this is the most usual difficulty in problem solving. People fail to make use of knowledge that they actually have. Certain constraints in people's minds, useful in other cases, prevent them from finding the way to the crucial idea.

Let me clarify the situation here. For a certain solver, all problems can be divided into two sets. The one set of problems includes the problems for which an algorithm is already in the mind. For example, if a linear differential equation is given, the same algorithm can be applied as was used in a similar case.

For the other set, the rest of the problems, only one general method can be applied: trial and error. If no known strategy seems to work, a trial and error method takes place in an attempt to reduce the difference between the givens and the goal. We apply anything available that seems likely to fill a part of the difference. The belief that we understand what this difference is guides us toward a certain subset of all possible alternative actions to try, thus reducing their number. So we often find a solution after some trials, perhaps after the first trial. However, there are many cases where this limitation of alternatives is the reason for failure. This is often the case, for example, when we get stuck at the start of a problem and realize, when given the solution later, that the relevant knowledge had been present all along.

Polya (1957) tried dramatically to help people in this difficult situation. Here is an excerpt from his instructions: "...combine the details differently, approach them from different sides...try to see a new meaning in each detail..." Similarly, Wickelgren (1974) tried to focus the reader's attention on the same matter, giving more specific instructions than Polya in the chapter entitled "Inference." Finally, I think that teachers, using special heuristic strategies, succeed in overcoming this difficulty following an indirect method, by "trapping" the problem into one of their strategies. But this succeeds only in a fraction of the cases, as I explained before. Besides, even if a problem can be solved by a certain known strategy, there are sometimes problems in knowing how to apply the strategy (Schoenfeld, 1980).

The instructions of Polya and Wickelgren are very good and have focused on the right issue, but they cannot always be applied. So, for those people who need more specific instructions, a problem arises: How can someone see new meaning in each detail, or how can someone force his mind to draw the right inferences with certainty? This matter has been bothering me for many years, and I have been trying hard to extend the instructions so that they would be more dynamic and active. As a teacher of mathematics for many years, I have had many opportunities to try out my new ideas on my students. After many trials, I have devised a very simple method which I call the "method of description." This method has proven effective in improving people's problem-solving abilities. The method tries to give power to people in very simple ways, so that they increase their chances to quickly see "the new meaning in each detail," or to draw the right inferences.

Here is the method, in its original form, as it is used in the classroom. First we read the problem many times until we have fully grasped the givens and the goal. We try for a while to solve it. Now suppose that the problem is not a routine one; none of the strategies we know seem to be applicable and no other idea comes to mind about how to solve it. The problem appears to be a new and difficult one. This is a case in which we would start applying our method. The method of description should be applied after having made such an investigation, or after having made attempts to solve the problem and having come to a dead end.

To use the method of description, we take the various parts or details of the problem and we try to recall all relevant knowledge. Because there is danger that some constraints may intervene and prevent the crucial idea from emerging, we should do this: While looking at a part of the problem, we temporarily try hard to forget the rest of the problem. We focus our attention on this part independently and then we start recalling anything relevant. We write down anything that we recall in short notes. This way of writing down short notes helps the mind in recalling more information, as well as the eye in recognizing quickly the required idea while scanning the notes later.

Some more analysis and clarification of the method will be given through examples. Three examples will be given: one puzzle which will be followed by further discussion and comments, and two problems taken from actual work in the classroom.

Application of the Method of Description: Example 1

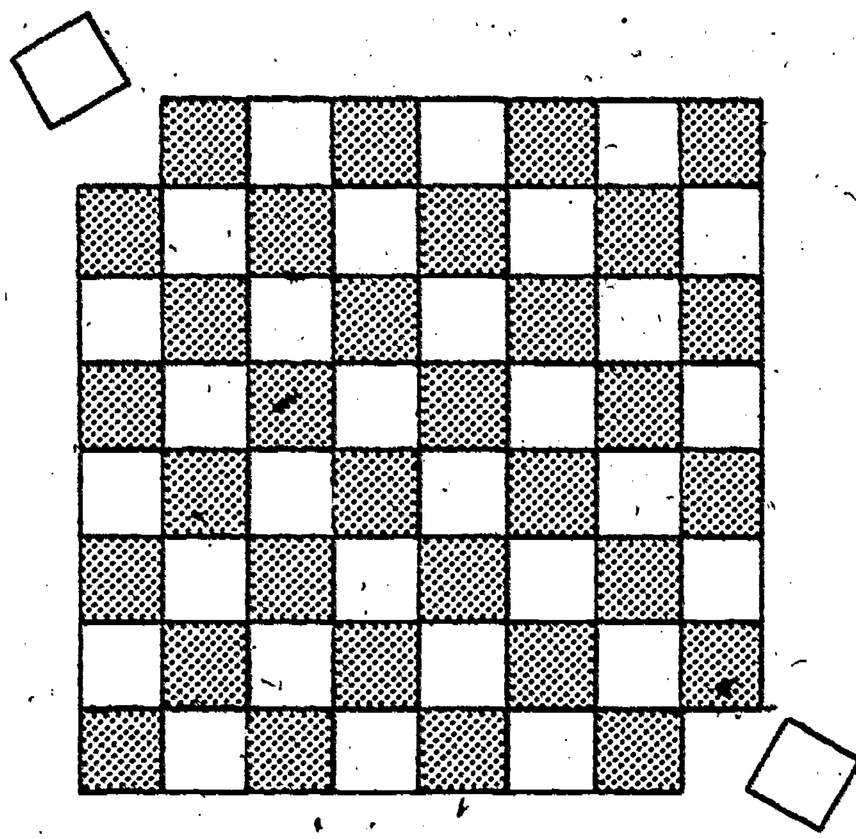


Figure 1

The problem. Consider the puzzle shown in Figure 1. You are given a checkerboard and 32 dominoes. Each domino covers exactly two adjacent squares on the board. Thus, the 32 dominoes can cover all 64 squares of the checkerboard. Now suppose one square is cut off from each of two diagonally opposite corners of the board. Is it possible to place 31 dominoes on the board so that all of the remaining 62 squares are covered? If so, show how it can be done; if not, prove it is impossible. You can try for a while to solve it before reading on.

Solution by the method of description. Many people find it very hard to solve this problem. Let us start the method of description. What do I see in this problem?

I have 62 squares.

I have 31 dominoes.

Each domino is rectangular. (feed back)

What are the properties of rectangles?

Parallel sides, congruent angles, congruent diagonals, etc.

What else do I see?

Let us focus on another detail.

Each domino covers two squares.

The two squares have a common side.

The two squares have different colors.

So one domino covers one black and one white square.

Here I see a simple correspondence.

1 domino covers 1 black square. and 1 white square. (feedback)

So, 31 dominoes cover 31 black squares and 31 white squares.
Is that possible?
How many squares do we have?
What are their colors?
Let us be careful here.
We have 62 squares.
Two white squares have been removed.
So there are 30 white and 32 black squares.

We can of course go on describing, but let us stop and take a quick look at everything we have written so far. Is it possible that the crucial idea is contained in what we have noticed? Yes, obviously. We have proved that 31 dominoes can cover 31 black and 31 white squares, but our checkerboard has only 30 white and 32 black squares, so it is impossible for the 31 dominoes to cover them all.

Discussion. What most people will have ignored or overlooked is the fact that both of the absent squares are white, and that each domino covers one white and one black square. This observation is the crucial idea, which unless you go about the problem in the proposed way (by describing in short notes), will probably escape you, as experiments have proven. The color of the squares is the information that most people fail to use. This information is in an implicit or semi-implicit form, and by describing we make it explicit.

Using the method of description means following this advice: If you cannot solve your problem, don't go straight to the target. Go around it. Describe everything fully. Include all the features that you notice (feedback). In a very short time the crucial idea will be written on the paper in front of you, and then all you have to do is to recognize it. It is not difficult to ignore the irrelevant data, which can be done by re-examining all the information you have recorded. If you have prepared your mind, as Pasteur said, it will be easier to recognize the crucial idea, and you will not become lost in the enlarged problem space. This is one of the reasons we use short notes.

Some readers may argue here that a danger of getting lost in the large problem space reduces the value of the method. But I can answer with the following: (a) Before starting to describe, we urge the student to "go around" the problem, etc., (a preliminary work). This is necessary, because then the mind will be prepared. Furthermore, searching around the problem to see whether any similarity exists between this problem and another with a known solution, or thinking if any known strategy can be applied, etc., can be considered to be a part of the description. After such an investigation we come to the detailed description, if we are already stuck. (b) One question: Which do you prefer, to abandon the attempt after getting stuck, to search in a narrow but sterile space, or to search in a somewhat larger space that most probably contains the crucial idea?

If you cannot find the crucial idea immediately, continue to draw more inferences by feeding back, describing, and scanning your data again. You must have patience, perseverance, a good understanding of the method, and the will to solve the problem. With practice, the usual hesitation to bring out the right idea will disappear. (Some strange inhibition or lack of courage prevents us from making information explicit that is presented in a subtle way.) These are the constraints I have already talked about.

It is rather like a battle. If we know only a few things about the enemy, it is unwise to go straight at him. It is better to go around the enemy first, to gather any information we can and to try to discover his weak points. In problems we have to find the subtle points.

There is something that I would like to emphasize here: A certain kind of training in the method of description has proven useful. Before giving students examples, they should be given training in which they are given little phrases and are asked to say as many things as they can. For example, give students the phrase "isosceles triangle" and ask them to describe just this phrase. The following might be part of the response:

It has two congruent sides.
It has two angles.
It is symmetrical about an axis.
The bisectors of the two equal angles are equal in length.
There are two congruent medians.
There are two congruent altitudes. (feedback)
The same is true of the medians, as well as of the bisectors.
(feedback)
What are the properties of congruent triangles?
etc.

If we train ourselves properly, we can extend this process almost infinitely. Ultimately we shall be able to do it silently, writing down only a few pieces of information, or not writing at all. If this method is applied by a class of students, we shall undoubtedly observe individual differences in terms of the inferences drawn. Where does this situation lead? First, there is the possibility that different ways of solving a problem will emerge. Second, if a problem is extremely difficult, the various ideas that have arisen will produce still more by feedback when presented to the class, and so our chances of hitting on the crucial idea are greatly increased.

Application of the Method of Description: Example 2

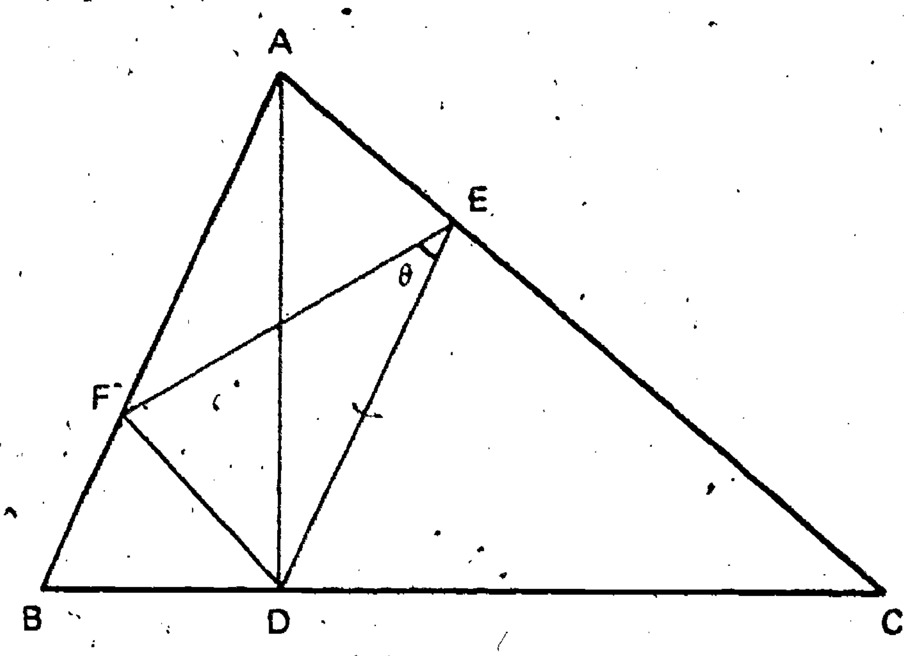
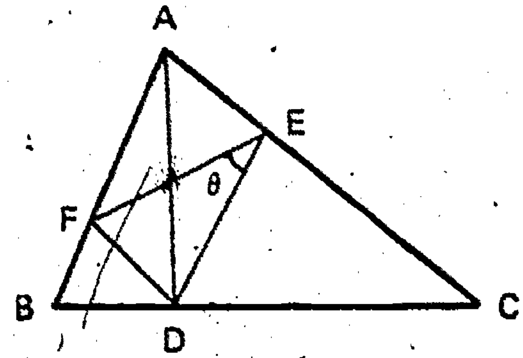


Figure 2

Another example involves a rather difficult problem of Euclidian geometry. I gave it once to a class of students, untrained in the method of description, as homework. The next day, they all came in complaining that it was too difficult to be solved. I also gave it to a class of students trained in the method of description. At first I reminded them of the method and I urged them to use it very intensively. A number of the students solved it. Here is the problem (see Figure 2) and the solution given by one student:

<u>The problem</u>. You are given an acute triangle ABC and its altitude AD. The goal is to construct an isosceles triangle EDF, (ED = EF) with angle DEF having a given measure θ, and the vertices E,F lying on the sides AC and AB correspondingly.

<u>Student's solution by the method of description</u>. Here is an example of the solution that a student gave, using the method of description:



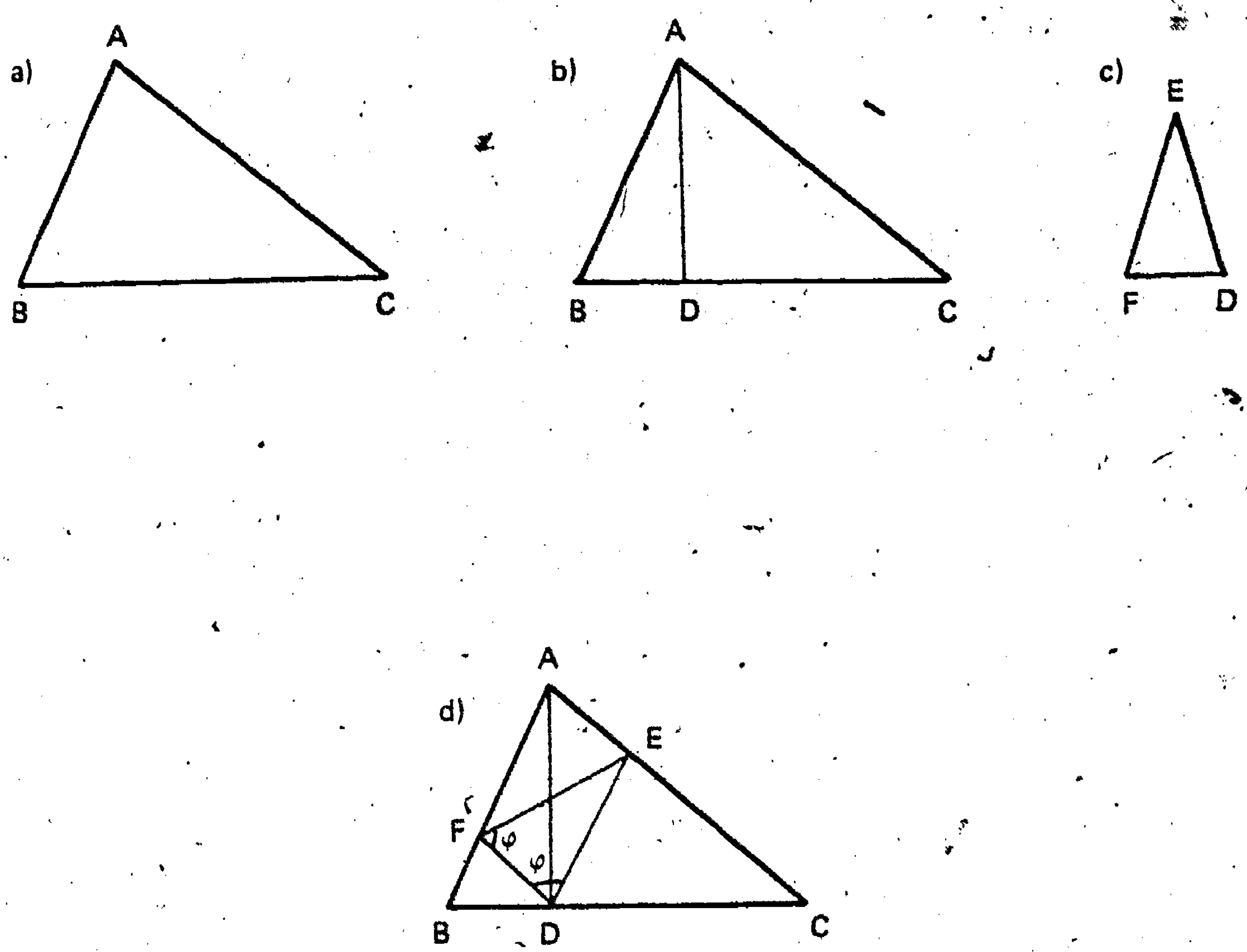


Figure 3

I can describe the main parts, but let me first separate them (see Figure 3).

1. Seeing ABC, what relations can I recall?

   Let me start writing.

   Angles A + B + C = 180 (A is the abbreviaition for m∠BAC.)

   $|BC - AC| < |AB| < BC + AC$ etc.

2. AD is the altitude.

   AD is perpendicular to BC.

   So we have right angles.

   We also have right triangles.

   Let me write any relations I recall:

   $(AB)^2 = (AD)^2 + (BD)^2$

   $(AC)^2 = (AD)^2 + (DC)^2$

   $(AB)^2 = (AC)^2 + (BC)^2 - 2(BC)(DC)$

   $(AC)^2 = (AB)^2 + (BC)^2 - 2(BC)(BD)$

   triangle ABD ⇒ D = 90 , A + B = 90

   etc.

3. Triangle EDF

   It is isosceles, ED = EF.

   m∠E = θ, m∠D = m∠F

   But m∠D + m∠F + m∠E = 180 (feedback).

   Solution of the equation 2(m∠D) + m∠E = 180⇒m∠D = (180−θ)/2

   So the measure of D is known: (180−θ)/2.

   That is, the angles of triangle EDF are constant.

   Triangle with constant angles. (feedback)

   The ratio of two of its sides is a constant: (DE/DF) = λ.[2]

   Also, (EF/FD) = λ, (DF/ED) = 1/λ

4. (See Figure 3d)

What do I have here now? (feedback)

m∠FDE is known, let us call it φ.

m∠DFE is also φ.

ratio (DE/DF) = λ, etc.

Can I find the position of E?

Can I exploit the above relations?

A known angle, a known ratio.

Yes, it reminds me of a known algorithm, the combination of a rotation and a dilation.[3]

---

[2]Let T be the set of all the isosceles triangles with angles θ and $(180-\theta)/2$, and let KLM be a certain triangle belonging to T. If K'L'M' is any other triangle of T, then KL/LM = K'L'/L'M'; that is, for any triangle, the ratio equals the constant KL/LM.

[3]In the students' geometry textbook, there was a chapter where various geometrical transformations were studied. One of the cases there was the case of a rotation about a point followed by a dilation. ((In our case, the rotation of AB about point D through the $(180-\theta)/2$ angle, and then the dilation of the image of AB, with center D and scale λ will map AB to another segment A'B'. The intersection of AC, A'B' gives the required E.))

Application of the Method of Description: Example 3.

The problem. Prove that $5555^{2222} + 2222^{2222}$ is divisible by 7.

Stop reading and try to solve it by yourself.

Two students' solutions. I once gave this problem to my students, who had been trained in the method of description. One of them began a description like this:

What do I see here?

I see the sum of two powers.

There appear the numbers 5555, 2222.

What inferences can I draw?

$5555 = 5 \times 1111,\ 2222 = 2 \times 1111.$ (feedback)

It reminds me of the properties of powers.

$(ab)^n = a^n b^n,\ a^{nk} = (a^n)^k$

$(5^2)^{1111} \times (1111^2)^{1111} + (2^5)^{1111} \times (1111^5)^{1111}$

or $(5^2 \times 1111^2)^{1111} + (2^5 \times 1111^5)^{1111}$

What do I see now? The sum of two powers with the same odd exponent 1111.

(feedback) It reminds me of the identity.

$a^n + b^n = (a + b)(a^{n-1} - a^{n-2}b + \ldots\ b^{n-1})$

So we have:

$(5^2 \times 1111^2 + 2^5 \times 1111^5)(5555^{2(1110)} + \ldots + 2222^{2(1110)})$

Stop reading and go on by yourself.

Let's now try to see what the remainder is of the division.

$(5^2 \times 1111^2 + 2^5 \times 1111^5)/7$

It is very easy now to find this remainder, either by making use of simple number theory, or by dividing the result of $5^2 \times 1111^2 + 2^5 \times 1111^5$, which is simple. So we see that the remainder is zero, and thus the problem is solved.

Another student made a different description:

What do I see here?

I see the sum of two powers.

I see 5's and 2's only.

What is the goal?

To prove that the given sum is divisible by 7.

Let me repeat:

5's, 2's, 7 ... 5, 2, 7.

Is there anything to connect these numbers?

$5 + 2 = 7$

or $2 = 7 - 5$

What about our problem?

$2222 = 7777 - 5555$

Let me use this:

$5555^{2222} + (7777 - 5555)^{5555}$ = [but according to the formula, $(a - b)^n = a^n - na^{n-1} b \dots - b^n$] $5555^{2222} + 7777^{5555} - \dots - 5555^{5555}$.

$7777^{5555} - \dots$ is obviously divisible by 7:

So the rest of the expression must be divisible by 7:

$5555^{2222} - 5555^{5555} = 5555^{2222} (1 - 5555^{3333})$

Since 5555^{2222} is not divisible by 7, $1 - 5555^{3333}$ must be divisible by 7, or $5555^{3333}/7$ must leave a remainder of 1. But since 5555 leaves a remainder of 1, so 5555^{1111} also gives a remainder of 1 (simple number theory).
Finally $1 - 5555^{3333}$ is divisible by 7 and the problem is solved.

The Method of Getting Out of Loops

I think this method will be better presented if we start immediately with examples.

Application of the Method of Getting Out of Loops: Example 1

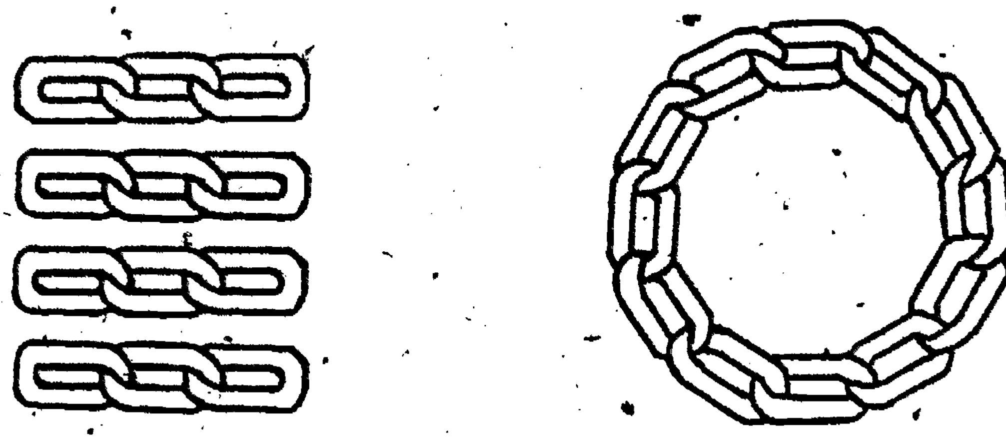


Figure 4

The problem. You are given four separate pieces of chain that are each three links in length (see Figure 4). It costs 2 cents to open a link and 3 cents to close a link. All links are closed at the beginning of the problem. Your goal is to join all 12 links of chain into a single circle at a cost of no more than 15 cents (see Figure 4).

Sample solution. Most people fall into a loop while trying to solve this problem. You can also try for a while to solve it, before reading on. Here is the method:

Divide the paper into two equal parts separated by a straight line. On the left half try to write a description of all the attempts you have made to solve the problem or all the procedures you have followed. If some attempts look similar don't take too long describing them all individually. It is useless. Describe them only according to their class. So it would be better to classify them before writing, i.e., suppose that you started by opening the end links of the first piece of chain and then you joined the one end of the first piece with one end of the second piece of chain. You continue by opening the other end link of the second piece of chain and joining it to one end of the third piece, etc. Continuing in this way you fail to solve the problem, so then you start again, for instance, by opening the end links of the second piece, and joining one of them with the end link of another piece, etc. Obviously the second procedure you have attempted is similar to the first, and if the first will not solve the problem, the second is unlikely to either. Always beware of similar approaches. After identifying such similarities, we go on describing on the left-hand side of the paper, while on the right-hand side we try our best to express all possible negations. One of these negations will often give us the crucial idea to enable us to get out of the loop. If you do not see the crucial idea immediately, go on and find some more negations.

A full description of attempts or procedures	All possible negations
What have I done? I have tried to join all the links into a circle.	
I have tried by opening the end links of one piece and joining them to the end of another piece, etc.	I don't open the end links, but I open other links. I open any links. I open the middle links. I open the end and middle links, etc.

Don't hesitate to write anything in the right-hand column which negates the first half. Now have a quick look at the right-hand side of the page. It is not difficult to recognize the crucial idea: Open the end and middle links. The solution is as follows: We open all three links of one piece - it costs 2 cents x 3 = 6 cents. Then we connect two end links with each opened link - 3 cents x 3 = 9 cents. The total is 6 cents + 9 cents = 15 cents.

If you still find it difficult to find negations, you can follow this simple way: Take one sentence from the left side, i.e., I have tried by opening the end links of one piece.

Now repeat the sentence inserting the word "not" in various places between the words. Some of the sentences produced in this way may have no meaning; however, some may have some meaning which is required. Let us try here:

I have not tried by opening the end, etc.

I have tried by not opening the end, etc.

I have tried by opening not the end, etc.

Obviously this last sentence includes the crucial idea of opening the middle link.

Discussion. Using this method you can succeed in getting rid of the strange hesitation people often have about changing procedure. Perhaps previously learned associations prevent our minds from grasping the correct procedure for a solution: Unnecessary constraints are often introduced that keep us on the wrong track.

We get stuck on the idea that we should join the chains end- to-end because that is the way in which we normally join lengths of similar objects, like pieces of string. We get used to working with certain ideas that we find convenient. For that reason, too much knowledge is often as much of a disadvantage as too little. When we have a great deal of knowledge, we sometimes get confused about which bits to use, and we are in danger of falling into the trap of getting stuck on one piece of information or one procedure. When we have only a little knowledge, we are more receptive to new ideas and methods. On the other hand, difficulties arise from the fact that we are unaware of the existence of similar problems or knowledge which might have helped us.

Application of the Method of Getting Out of Loops: Example 2

The problem. You are given six matches and you are asked to make four equilateral triangles with them.

Try for a while before reading on.

A sample solution using the method of getting out of loops.

A full description of attempts or procedures	All possible negations
First I made one triangle	I don't make only one triangle first, but more.
Then I tried to make a second with a common side.	I don't try to make a second, but two more.
This left only one match. Is there anything I have not mentioned?	
I am trying to do it on my table, etc.	Let me not try doing it on a table. Let me try doing it not on a plane. Let me try doing it in space. (crucial idea)

The solution is to make a tetrahedron in three dimensions (see Figure 5).

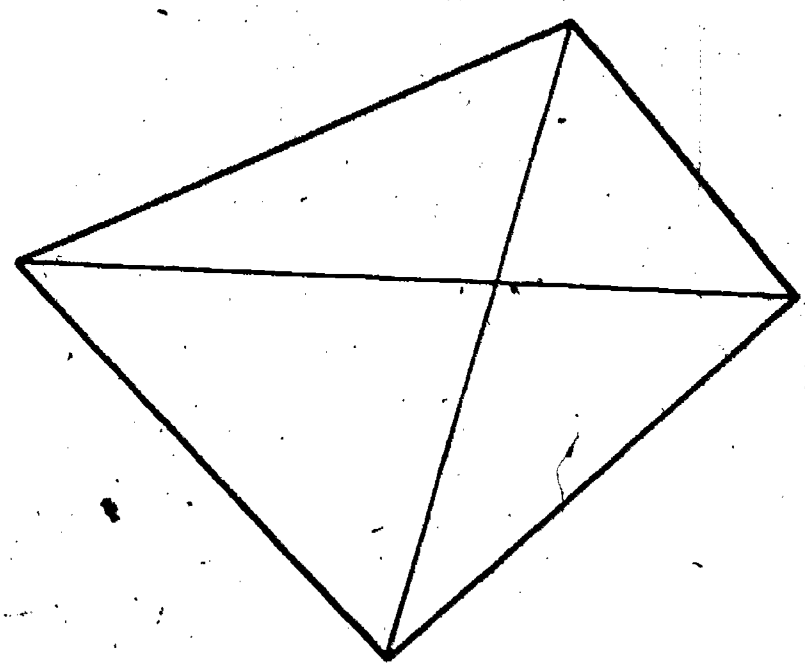


Figure 5

Discussion. After some practice we shall be able to go through this process without going through all the steps. We shall be able to find the right idea or the right procedure quickly, by writing down only a few pieces of information. This method can be summarized as follows: Analyze fully what you have done until now, then try to express negations in all possible ways. This increases flexibility and helps you to change procedures after you have gotten rid of all unnecessary constraints. Maybe here it is the word "triangle," which refers to a figure in a plane, and which constrains our brain to attempt solutions in a plane.

Application of The Method of Getting Out of Loops: Example 3

The problem. Eliminate "a" from these two equations:

$$x^3 \sin a + y^3 \cos a = c^3 \sin a \cos a$$

$$x^3 \cos a + y^3 \sin a = c^3 \cos 2a$$

Try for while before reading on.

A sample solution. If you try to eliminate "a" by expressing "a" or sina or cosa in terms of x and y, you will come to a dead end. Go on applying our method by yourself.

Procedures	Negations
I try to express sina in terms terms of x,y.	Let me not try to express sina in terms of x,y.
I try to express cosa in terms of x,y, but the same difficulty arises.	Let me not try with cosa either.
I try combinations by trial and error, but again I come to a dead end.	Let me try another transformation. Let me try to express not sina. Let me express x and y in terms of sina, cosa.

We regard these two equations as simultaneous equations in which x^3, y^3 are the variables. A similar problem would be:

$$ax^3 + by^3 = c$$

$$a'x^3 + b'y^3 = c'$$

Why should we follow this procedure? Never ask this question in problems. At this point we cannot know where the procedure will lead, or, indeed, if it will produce anything useful. This question will lead to dangerous hesitation. We always hope that the new form of the problem will show us the way to the solution, and this very often happens. After solving this system, we simplify and finally:

$$\begin{matrix} x^3 = c^3 \cos^3 a \\ y^3 = c^3 \sin^3 a \end{matrix} \Rightarrow \begin{matrix} x = c \cos a \\ y = c \sin a \end{matrix}$$

Thus we arrive at two very simple equations, and the solution to our problem is obvious. I can solve the statement $\cos a = x/c$ and substitute the second $y = c \sin a = \pm c\sqrt{1 - \cos^2 a}$, or better:

$$y^2 = c^2 (1 - \cos^2 a) \text{ and thence}$$

$$y^2 = c^2 ((1 - (x^2/c^2))$$

Discussion. In this problem, many people introduce the unnecessary constraint of trying to express the variables to be eliminated in terms of the others, because in other problems they have followed this procedure. Try to get away from the preconceived ideas you have in your head. So, if you feel you have gotten into a loop:

Negate your progress to date.

What other possible courses exist?

Negate the method you have been using.

Negate the form of the algebraic statement in the problem, if there is one, and then continue. That is, make all possible transformations.

Refuse to look at a form from one particular angle.
Look at it from others, too. Take auxiliary lines - then more auxiliary lines.
Negate the angle from which you have approached the problem.
Describe more details.
Negate your description to date.
What other description could have taken place?
Where else could you have started?
Give the whole thing another interpretation. Re-interpret each detail.
See what other people have to say. Talk about the problem with other people.
If you are tired, leave the problem for a while.
Leave it for longer. Do something else for a while.
If it is late at night, leave it and go to bed. Perhaps in the morning you will have an idea, or a fresh course will occur to you.
If you are taking an examination, go on to another problem.
Come back to it later. Perhaps you will now be able to see a new approach.

To recapitulate: If you get stuck with a problem, of whatever kind; if your research project bogs down; if you feel unable to press forward despite the description you have made, then I repeat, make a careful analysis of what you have done. Classify your actions into similar groups, if there are any. And finally, make all possible negations. One of them will show the correct way to the solution. But if it does not, don't give up. Carry on and find more negations.

A Solution Process Based on the Method of Description

This section presents a more specific version of the method of description in the form of an informal specification of a problem-solving process. A goal of this formulation was to provide a basis for formulating the method explicitly, as a computer program.

Several slightly different lists of component processes could be made. The method of description refers to how we can find the crucial idea, not to how to produce the whole problem solution. However, with the addition of a few more components, our list of processes can give the whole solution. We can decide about these few extra things, so that our list could better be used with certain classes of problems, for example, problems to prove. We can make a somewhat longer list if we want to fit more classes of problems, but in any case, the method of description included in the list is essentailly the same. Here I shall present such a list, which is most suitable for proof problems in pure mathematics.

A List of Component Processes

1. Describe of the whole problem through your known method of solving: Check whether some method known to you can be applied (contradiction, backwards, subgoals, other strategies). If you failed, go to (2). If some of your known methods seem to be applicable, but you come to a dead end after trying them, you can go to (2), then (3), and then (1) again. If you still fail, follow (6), then (7), then (1) again. A more powerful trial is to go through (3), (6), (7), (12), and then (1) again.

2. Describe the givens: Without introducing new elements, enrich the problem space with the goals of any theorems having the same givens. (Elements in geometry are the points, segments, etc.; elements in algebra are the variables, etc.)

3. Describe the main parts of the givens: Enrich the problem space with the goals of any theorems that have each of these parts as their givens. If you apply the contradiction method, attach the negation of the goal to the givens.

4. Now look at the enriched problem space: Is the goal of your problem included there? If yes, the solution is found.

5. If (4) does not succeed, then do this: Find all theorems with the same goal as your problem's goal (description of the goal), put them in random order, and examine carefully whether the givens of any of these theorems are included in the enriched problem space. If yes, then the initial givens and the chain of theorems used to reach all the givens of our theorem, together with the goal, are the solution. Eliminate the rest of the elements.

6. If (5) does not succeed, then do this: Take each subset of the problem space separately and describe as in principle (3). Now go to (7).

7. At the new problem space apply process (6). This is enrichment by feedback. Go to (8).

8. Try process (4).

9. If (8) fails, try process (5).

10. If (9) fails, check why it failed. Find what was missing, a property, an element, or both. If only a property is missing (for example, the congruence of two angles), follow once more the previous procedure (7), (9). If element(s), or elements and properties are missing, find the case in which their number is least. Take this case (out of other theorems, if there are more than one theorem having the same goal as ours) and add the missing elements according to the theory. If there are alternatives to the missing elements, choose the one closest to the givens. If there still is some property missing, in the new problem space, apply process (6), then (5). If no property is missing, the solution is found (unusual case).

11. If there is not a backward step from the goal, so (5) fails from the beginning, try to find an equivalent form of it, and then try the previous procedure again, (5), (6), etc.

12. If you still fail, then do this: Divide the initial givens and the parts of the goal into subsets containing at least two things (elements), or elements and properties. Check to see whether there is any theorem with any of these subsets as givens. Apply its goals, even if it introduces new elements. Now follow again the previous procedure, (1).......(11) of description.

Discussion

I hope that it has already become obvious that our list tries to exploit to a maximum degree the solver's knowledge of problem solving. We also consider that the solver has a good idea of some methods of contradiction, working backwards, subgoals, and also that he knows when to apply them (see Wickelgren, 1974). Trying to exploit all the existing knowledge is the essence of the method of description. If some points of our list are unclear, the following section may help you because a flowchart will be presented there using the list.

A Computer Simulation of the Method of Description

Here a computer program will be described which solves problems using the method of description. Some alternatives at the level of detailed implementation that can be used will also be described. We have programmed the model as a production system written in ACTP. The details of the program will not be presented here, but only an outline of it, and its application in two rather difficult problems. The application will be based mainly on our list of principles, 1 - 12.

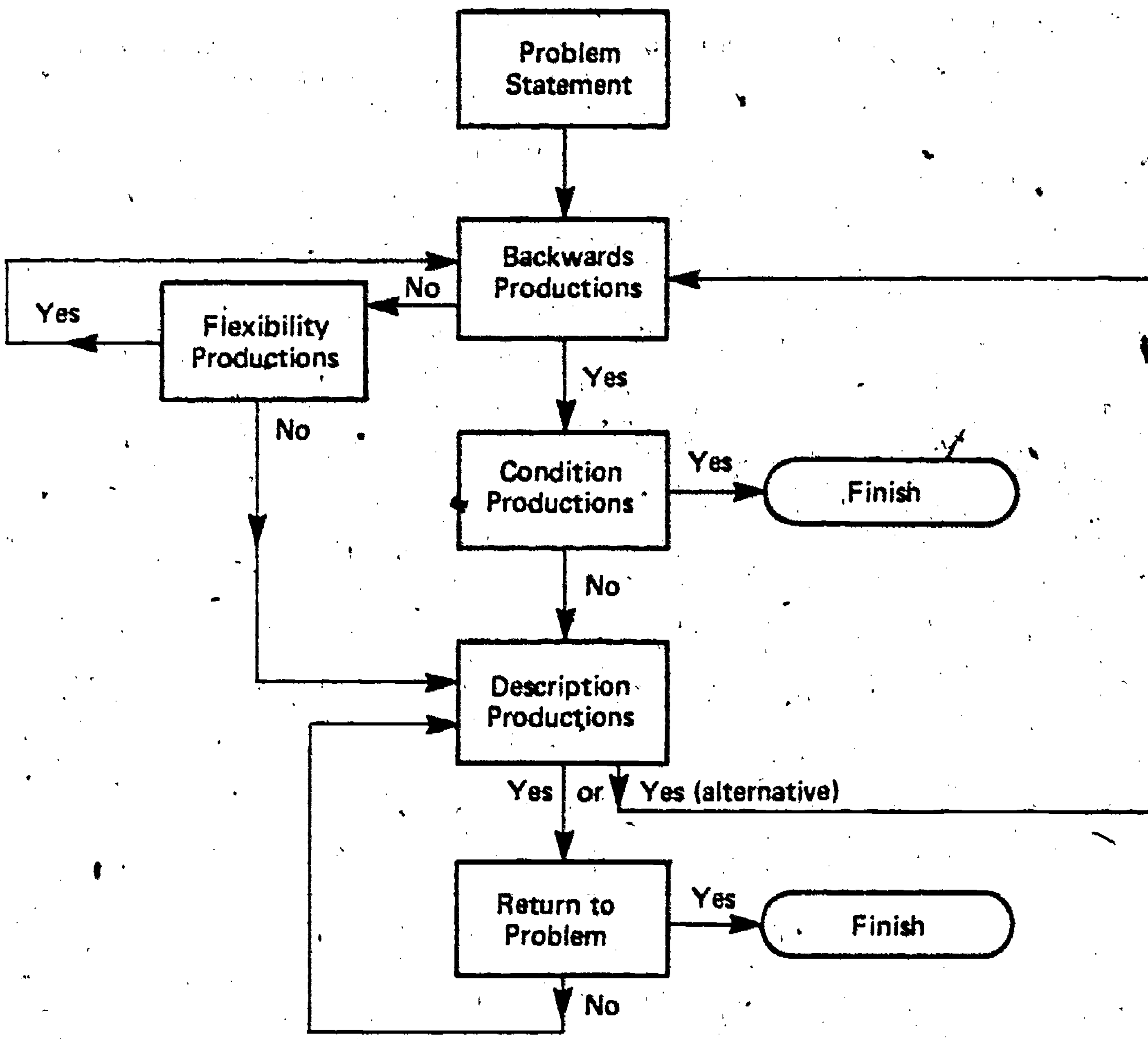


Figure 6

Figure 6 shows the fundamental ideas of the program in the form of a flow chart. I will discuss each component of Figure 6 briefly.

The "Problem Statement" is represented as a data structure, containing the givens and the goal of the problem.

The problem-solving knowledge in the program is represented in the form of productions. First I will discuss "Backwards Productions." These are productions corresponding to the goals of the theorems. Each of these productions tries to match the goal of a problem to the goal of one of the theorems. If such a matching succeeds, then the program comes to the "Condition Productions." If not, it comes to the "Flexibility Productions." If it comes to the "Condition Productions," it does "Description of the Goal." If it finds the theorem or the theorems that correspond to the goal found in the "Backwards Productions," then it tries to match the givens of the problem to the givens of any of these theorems. If such a matching succeeds, then the solution is found. If not, then the program comes to the "Description Productions."

The "Description Productions" are the most important part of the program because here the real description takes place, enriching the givens by a forward procedure. The following strategy is used: First, the program takes all the theorems, and divides them into two classes. The one class contains theorems that simply transform a statement into an equivalent one, or build relations without introducing new elements. For example, the associative law, which transforms a + (b + c) into (a + b) + c, or the property of the identity element, which here transforms a,0 into a + 0 = a. The second class contains theorems that introduce new elements. For example: two points A,B give a segment $\overline{AB}$, or if "a" is our element, then there is "-a" which is the inverse element, a + (-a) = 0.

The program uses the first class of productions to produce new statements without making dangerous enrichments of the problem space by introducing new elements. After the theorems on the first list are applied to the givens of our problem, the process returns to the "Backwards Productions". If no solution is found in the new enriched problem space, the program makes one more description by feedback and goes to the beginning again.

If there still is no solution, then new elements are introduced economically, one by one. When a new element is introduced, the program goes back to the first list, etc. The number of times the program will run the first list before coming to the second can be decided by the user. Details such as use of a new element right from the beginning, or the order of theorems in the lists also can be changed to vary the programs's performance. Such changes would sometimes change the speed of solution, because they would speed up the enrichment of the problem space. (The feedback procedure is unchanged by such variations.)

The program has the possibility of finding different solutions to the same problem. For example, if in the "Conditions Production" there are two or three theorems with the same goal, and if it happened that we found a solution by matching the first theorem in this list, a different solution could be obtained by changing the order of those theorems. Thus, the program can be used to generate a set of possible solutions to the problem relative to the set of theorems, or it can warn us if the problem cannot be solved with these theorems.

Two alternative processes involve the action that follows the "Description Productions." Instead of coming to "Backwards Productions," the process could come to another production, a single one called "Return," where there are only two conditions. The condition contains the whole statement of our problem, the givens and the goal. If the program finds them both in the enriched space after the description, then the solution is found. This is clarified in the following example: If we want to prove that in a ring we have a x 0 = 0, we take the givens and apply description. Then coming to the enriched problem space, the program checks to see whether such a relation, a x 0 = 0, exists. This process also can be used if we have a problem in geometry, where we want to prove, for example, that two segments are equal, a = b. In the enlarged "givens" space, there may be many such equalities of segments. That is why the whole problem is included in the test that is made in the process "Return." Otherwise, the program could not distinguish valid solutions from any arbitrary relation of equality.

If in "Backwards Productions," there is no such goal to match, then the system comes to the "Flexibility Productions" where it transforms the goal into an equivalent one by applying operations. Then the program returns to the "Backwards Productions." If this does not succeed, then the program comes to the "Description Productions," and follows the straight-forward procedure. When the "Description" process has been preceded by the "Flexibility" process, the program uses the "Return Production." If the "Return Production" fails, the program comes back to the "Description Productions," etc.

Application of the Computer Program: Example 1

The problem. Prove that $a \times 0 = 0$.

The program's solution of the problem. The program was given the following inference rules, in the form of productions:

1. if a is an element, then $a = a$
2. if $a = b \Rightarrow b = a$
3. if $a = b$, $b = c \Rightarrow a = c$
4. $a + b = b + a$
5. $ab = ba$
6. $a + (b + c) = (a + b) + c$
7. $a(bc) = (ab)c$
8. $a + 0 = a$
9. $a + Q = a \Rightarrow Q = 0$
10. $a + (-a) = 0$
11. If $a = b$, $c = d \Rightarrow ac = bd$, $a + c = b + d$
12. $a(b + c) = ab + ac$
13. a,b el $\Rightarrow a \times b = c$ is element, $a + b = d$ is element
14. if a is an element, there is $-a$: $a + (-a) = 0$
15. Substitution law

Properties (2), (4), (5) were automatically applied by the program.

Properties (6), (7), (10), (12) were given in this way: If you see the first side anywhere, you can replace it by the second, i.e., if you see $a(b + c) = d \Rightarrow ab + ac = d$.

Property (9) has the meaning of the uniqueness of 0, which was also proved by our program before. (It is not included the proof given below.)

Properties (13) and (14) were programmed not to be used at the beginning, but after a repeated failure. (In order to avoid introducing new elements which might dangerously enlarge the problem space, they are the second list of the "Description Productions".)

The program was run in three different ways. Presented here will be one of them, which was done in 21 steps:

1. We have the given.

2. We have the goal.

3. The program looks at the list of the goals, and it cannot find any goal like $a \times 0 = 0$ (Backwards procedure).

4. The program decides to transform (a little?) the goal to an equivalent one, so it takes the first side, goes through the theorems and sees that it can match with (13). So it transforms the goal $a \times 0 = 0$ into $v_2 = 0$ where $v_2 = a \times 0$.

5. The program now comes back to the list of goals and tries to match any. Then it sees that it can match the goal of Theorem (9).

6. Next the program looks for the condition of the Theorem (9), that is, for $a + v_2 = a$ in the problem space.

7. The program fails to find this, so it starts describing.

8. The program tries the reflexive law, sees that that it can be applied, and it does so, $v_1 = v_1$.

9. Since v_1 and 0 exist and property (8) can be applied, the program builds the relation $v_1 + 0 = v_1$.

10. The program fails at Theorems (6), (7), (8).

11. The program fails at Theorems (9), (10).

12. The program succeeds in Theorem (11) because here we have two equalities $v_1 = v_1$ annd $v_1 + 0 = v_1$. Here the relation, $v_1 (v_1 + 0) = v_1 v_1$, is built.

13. The distributive law is applied: $v_1 v_1 + v_1 0 = v_1 v_1$.

14. The program checks whether the proof is found, by looking at the enriched space for both relations: $a + v_1 0 = a$, $v_1 0 = 0$.

15. It fails, and starts again describing from the beginning by feedback (6). Although we had decided not to introduce new elements, in order to avoid dangerous enlargements, a new element was mistakenly introduced: $v_1 v_1 + v_1 0$ was named as a new element. Fortunately enlargement was not too large! So here the reflexive law is applied to the new element $v_1 v_1 + v_1 0 = v_1 v_1 + v_1 0$.

16. Theorem (8) is applied: $v_1 v_1 + v_1 0 + 0 = v_1 v_1 + v_1 0$.

17. Theorems fail (6), (7), etc.

18. $v_1 v_1 + v_1 0 = v_1 v_1 \Rightarrow v_1 0 = 0$, Theorem (9) is applied.

19. It assures that Theorem (11) is already applied.

20. It assures that the distributive law is already applied.

21. It checks and finds both the givens, $(v_1, 0)$, and the goal, $(v_1 \times 0 = 0)$, in the enriched problem space. So it reports "Finish," that proof is found.

We made two alternative versions of the above program. In the first we changed the order of the theorems, so that (9) was put after (12). Here we also made the program go to the "Backwards Productions" after description. In this way, the solution was found sooner. The second alternative used a forward procedure by describing (eliminating "Backwards") and again the solution was found sooner.

Here is an outline of the solution of the problem to prove $a \times 0 = 0$, following failure of an attempt to solve the problem by backward search.[4]

1. $a = a$ description (reflexive law)

2. $a + 0 = a$ description (identity element property)

---

[4]Often in the past I have given this problem to students of mine. They found it extremely difficult to solve. They could not make the combinations (1) - (2) - (3) - (4).

3. $a(a + 0) = a \times a$ description by feedback (uniqueness of operation)

4. $(a \times a) + (a \times 0) = a \times a$ description by feedback (distributive law)

   Trials-fails

5. $(a \times a) + (a \times 0) = (a \times a) + (a \times 0)$ description by feedback (reflexive law, useless)

6. $(a \times a) + (a \times 0) + 0 = (a \times a) + (a \times 0)$ description by feedback (identity element, useless)

7. $(a \times a) + (a \times 0) = a \times a \Rightarrow a \times 0 = 0$ (uniqueness of identity element)

8. Proof is made.

<u>Application of the Computer Program:</u> <u>Example 2</u>

A short version of the solution for a geometry problem will be presented here, as it was done by the program. The usual theory of geometry was used.

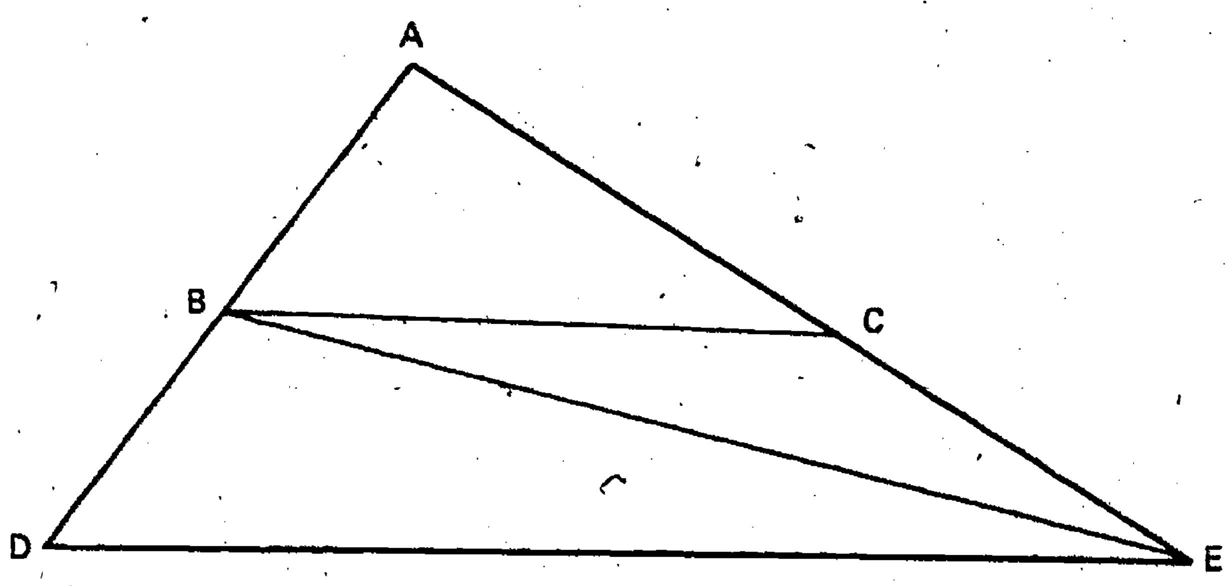


Figure 7

<u>The problem</u>. A scalene triangle ABC is given. We extend AB and AC so that BD = CE, and we draw the segment DE. The goal is to prove that DE > BC (see Figure 7).

<u>The program's solution of the problem</u>.

1. The program goes to the Backwards Productions, and tries to find any theorem with the same goal. Three such theorems should be found.

   a. If ABC, A'B'C' are triangles and AB = A'B', AC = A'C', $\angle A > \angle A'$, then BC > B'C'.

   b. If ABC is a triangle and $\angle B > \angle C$, then AC > AB.

   c. If AB > DC and DC > EF, then AB > EF.

2. Then the program comes to the next list of productions, and tries to see if the givens of any of the 1a, 1b, 1c are included in the givens of our problem. This fails and so description starts.

3. In this case, the program was run with introduction of an auxilliary element from the beginning. So, while the program applied the definition: If P,Q are points, then PQ is a segment, it built up the segment BE. In the next productions of the description, the program applied all the theorems of geometry that were applicable so:

4. It noted the existence of the triangles ABC, ABE, BCE, BDE, ADE.

5. It noted their sides, their angles.

6. It noted the external angles of triangles ABC, ABE, BCE.

7. It found relations between external angles and internal ones, like:

   $\angle DBC > \angle BAC$

   $\angle DBC > \angle ACB$

   $\angle DBE > \angle BAE$

   $\angle DBE > \angle AEB$, etc.

8. The program noted congruence between the angles such as $\angle BEC = \angle BEA$, $\angle EDB = \angle EDA$, etc.

9. The program applied the substitution law in (6) and (7), so it yielded $\angle DBE > \angle CEB$.

10. After this, the program came back to the backwards procedure. It found that a matching occurs with case 1a, and reported "Finish," that is, proof is found.

An alternative to this was run, where instead of going back to "Backwards Productions" after the description, it came back to "Return." Here the situation givens-goal is matched to the enriched problem space (as in the previous example).

Discussion. In the "Description Productions," the order of the theorems was such that a quick solution was produced. Otherwise, the solution would be found after the second or at most the third use of "Description" by feedback description. We also omitted a large number of theorems in the productions, because they would not apply in this problem (so it would just be a pain for us to write them down, for example, the theorems on parallels). But, if in spite of this, we had written down all the theory of geometry in the form of the production system we refer to here, a quite feasible task, then it is obvious that the description of the geometry problem included in the Method of Description section, could have been made very easily by our program, perhaps more easily than it was made by the students.

Conclusion

Finally, some general requirements that should be fulfilled by a good problem solver, either a computer program or a human solver, are:

1. In the LTM, there should be all the required knowledge for the problem (axioms, definitions, theorems, etc.). As you cannot produce energy from zero, so you cannot produce a new piece of knowledge from none. One could argue here that some of the required knowledge can be built up by combining the previous theorems, axioms, etc. For instance, if there is a chain of 40 theorems in an algebra textbook, and we ask somebody to prove the fortieth theorem by using only the first twenty ones, one could argue that the required knowledge exists, since the omitted theorems can be proved, etc. But since, as we have already said, every solution includes, more or less, some trial and error, the probability of succeeding in a very long chain of successful trials is very small, if there are time limits. (In the history of science, where no time limits exist, months, years, or even centuries have sometimes been needed.)

2. The description method, as we have it here. The solver must be able to apply the method easily.

3. The method of getting out of loops. The solver must be able to apply this method easily, too.

4. Heuristic strategies are always welcome, because they can speed up the way to the solution. They should be placed in the first step of the description method, and they can be well adapted to our computer program, as a part of the theory. They are more valuable in practical matters, such as in the usual examinations in schools, universities, etc., but are less useful in real research, in original problems, and in domains other than the ones to which they refer. For example, algebra strategies are almost useless for physics.

5. The method of subgoals. The solver should be able to decide at the very first steps of the description whether this powerful method can be used.

6. The working backwards method. The solver should know when and how to use it (see Wickelgren, 1974).

7. The contradiction method. The solver, here also, should know when and how to use it (see Wickelgren, 1974).

The three methods, (5), (6), and (7), could be included in the computer program that we already presented, after a partial differentiation of the production. We have already used the contradiction method in solving two problems.

Acknowledgements

Here I would like to say a few words about the three people who have most influenced my ideas about problem solving. First for G. Polya, I have to say that he is the pioneering father of problem solving, whose books are an everlasting source of inspiration and enjoyment for me. Second, for W. Wickelgren, I want to emphasize this: Some 5 or 6 notes from his "How to solve problems" struck me and fertilized other ideas of mine, as they seemed to be on the same "wavelength," so that after a time, the ideas of this paper were born. Third, for Jim Greeno, the only one I know personally, I have to say that he was just the right person at this time: He gave me the right orientation for the further study of my methods, so I was able to find their framework and consequently to make them more steady and rigorous. I wish to thank him here for the immense amount of help he gave me.

S. Kalomitsines

Dec. 4, 1980

References

Kalomitsines, S. P. (1978). ΜΑΘΕΤΕ ΠΩΣ ΝΑ ΛΥΝΕΤΕ ΠΡΟΒΛΗΜΑΤΑ (Learn how to solve problems). Athens: Anastasakis.

Kalomitsines, S .P. (1980). Attack your problem. Athens: G. Tsiveriotis.

Newell, A. & Simon, H. (1972) Human problem solving. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Polya, G. (1980). How to solve it. New York: Doubleday and Company.

Polya, G. (1962). Mathematical discovery. Volume 1. On understanding, learning and teaching problem solving. New York: John Wiley and Sons.

Rubenstein, M. (1975). Patterns of problem solving. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Schoenfeld, A. (1980). Teaching problem-solving skills, American Mathematical Monthly, 87, (10) 794 - 805.

Wickelgren, W. A. (1979). Cognitive psychology. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Wickelgren, W. A. (1974). How to solve problems. San Francisco: W. H. Freeman Company.